转换概率和词长期待对语音统计学习的影响*

于文勃[1] 王璐[1] 瞿邢芳[1] 王天琳[2] 张晶晶[3] 梁丹丹[1]

（[1]南京师范大学文学院，南京 210097）（[2]纽约州立大学奥尔巴尼分校教育学院，纽约 12222）

（[3]南京师范大学心理学院，南京 210097）

摘要 语音统计学习指个体在加工人工语言过程中，可以追踪音节间的转换概率实现切分语流、提取词（语）的过程。本研究采用 2（转换概率：高转换概率、低转换概率）×2（词长期待：两音节、三音节）的混合实验设计来考察转换概率和词长期待对语音统计学习的影响，转换概率是被试间变量，词长期待是被试内变量。事后检验发现, 仅在低转换概率人工语言的三音节迫选条件下，被试没有表现出显著的学习效果。事先对比发现，在学习低转换概率的人工语言后，被试完成三音节迫选试次的成绩显著低于两音节迫选试次；在三音节迫选试次中，学习低转换概率人工语言被试的成绩也显著低于学习高转换概率被试的成绩。以上结果说明, 转换概率和词长期待共同影响个体语音统计学习的效果。

关键词 语音统计学习，转换概率，词长期待，韵律

1 前言

统计学习机制指个体能够从外界输入的时间信息和空间信息中发现概率规律并以此规律学习新事物（Frost et al.，2020；Batterink et al.，2015；Bogaerts et a.，2016；Siegelman，Bogaerts，& Kronenfeld et al.，2018；Saffran et al.，1996；Saffran & Kirkham，2018；Frost et al.，2015），大量实验已经证明，成人、儿童甚至新生儿都具有此项“天赋”（Wang & Saffran，2014；Bosseler et al.，2016；Kidd & Arciuli，2016；于文勃，梁丹丹，2018）。统计学习中的核心概念是转换概率（transitional probabilities，TPs），指的是相邻出现元素间的概率关系。比如，在词组“pretty boy”中，音节“pre”和“tty”是词内音节，前一个音节出现后，后一个音节出现的几率较大，因此转换概率较高；相反，音节“tty”和“boy”是词间音节，前一个音节出现后，后一个音节还可能是“girl”、“dog”等，因

收稿时间：2020-7-11

于文勃和王璐为本文的共同一作。

* 江苏省社会科学基金项目成果(19YYC003)；江苏高校优势学科建设工程资助项目(PAPD)。

通信作者：梁丹丹，E-mail: ldd233@163.com

此转换概率较低。研究者普遍认为个体将转换概率较低的位置视为语流中的词边界，并且将转换概率较高的音节组合作为词的语音表征存储在记忆中（Erickson et al.，2014；Estes et al.，2007），这类研究也被称作语音统计学习或概率词切分[1]。

在语音统计学习研究中，Saffran 等（1996）首次考察个体利用概率信息切分语流的过程，如图 1，作者使用 12 个音节合成四个三音节词（每个大写字母代表一个音节），随后以三音节词为单位按照伪随机的方式合成人工语言，不仅要求每个词不能连续出现两次，还在合成人工语言过程中避免了重音、停顿等韵律信息。实验采用“学习-再认”范式[2]，先向 8 个月大的婴儿呈现 2 分钟的人工语言，随后分别向他们呈现目标词（合成人工语言的原始三音节词，如 ABC）和跨界词[3]（由前一个目标词后两个音节和相邻后一个目标词第一个音节组成的词，如 BCD，EFA）。结果发现，婴儿对跨界词注意时间更长，对目标词注意时间更短，表现出新奇效应。作者推断，目标词 ABC 后面可能出现 DEF、GHI 或 JKL，音节 C 后面接任意一个音节的转换概率均为 1/3，明显低于目标词内音节的转换概率（为 1），因而婴儿会在音节 C 后进行切分，将音节组 ABC 作为词储存在记忆中，但由于跨界词没有被切分出来，婴儿会表现出新奇效应。

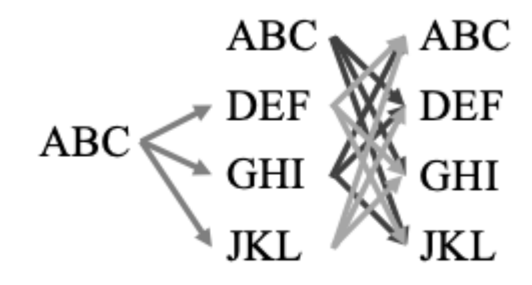


图 1 语音统计学习任务人工语言合成规则示意图

根据统计学习的理论框架，转换概率越高，音节之间的连结关系越紧密，个体对它们的记忆会更加牢固，即学习效果会更好。假设 A、B 两种人工语言中仅有目标词的转换概率有所区别，在人工语言 A 中，目标词的转换概率为 1，在人工语言 B 中，目标词的转换概率为 0.6，两种人工语言中非词的转换概率均为 0。如果转换概率是影响统计学习效果的核心因素，同时人工语言 A 中目标词和非词之间转换概率的差异更大，那么，

---

[1] 统计学习任务可以分为听觉和视觉两种形式，但在听觉模态下还包括音调（Saffran, 2010）、声音（Siegelman ，Bogaerts，Elazar et al.，2018）等亚类的统计学习研究，本研究考察的是以音节为载体的统计学习机制，学者普遍以语音统计学习（verbal statistical learning）来进行指代。

[2] 在婴儿实验中常使用转头偏好任务，成人实验中常用使用二选一的迫选任务，两种任务本质上均为再认范式，由于本研究被试为成人，以迫选任务作为测验任务。

[3] 统计学习研究中通常会有两类干扰刺激，一类是此处描述的跨界词（partwords），另一类被称作非词（nonwords），指的是词内部音节没有相互连接的可能性，比如 AEL，这类词内部的转换概率为 0。

被试切分人工语言 A 的效果应该比切分人工语言 B 的效果更好。虽然大量研究都证实了个体可以通过目标词和跨界词转换概率的对比关系切分语流，（Estes , Gluck，& Bastos，2015；Estes & Lew-Williams，2015；Palmer & Mattys，2016；Potter et al.，2017），但尚未有研究考察当目标词的转换概率发生变化时，被试完成迫选测验的成绩是否会发生变化。对这一问题进行考察可以与现有相关研究形成互补，进一步证明个体确实是通过计算音节间的转换概率来切分连续语流，并同时证明转换概率具有心理现实性。目前，仅有一篇视觉统计学习研究考察当目标词或结构的转换概率变化时，个体的切分效果有无变化。Bogaerts 等（2016）在实验中设置了 0.6、0.8 和 1.0 三种转换概率强度，结果发现，当转换概率从 0.6 提升到 0.8 时，被试的正确率仅有小幅度的提高，但从 0.8 提升到 1.0 时，被试的正确率有了显著提高。虽然视觉统计学习和语音统计学习都认可转换概率对学习效果的决定作用，但一些研究发现两种模态下的统计学习结果相关较低（Siegelman & Frost，2015），对于个体发展来说，视觉统计学习能力随着年龄逐渐提高，而听觉形式下则没有明显变化（Arciuli & Simpson，2011；Raviv & Arnon，2017）。Emberson 等（2019）也发现，对于同一年龄段婴儿来说，听觉统计学习能力要强于视觉统计学习能力，听觉模态下的统计学习能力可能要发展得更早、更快。从以上分析来看，难以将视觉统计学习结果直接推广到听觉模态，而在听觉模态下考察转换概率对统计学习效果的影响，不仅有助于探究这一模态下统计学习过程的特点，也有助于在一般领域下揭示其普遍性和独特性。

研究者普遍认为，个体可以通过计算音节间的转换概率来实现统计学习过程，因此，转换概率因素可以看作是自下而上形式的信息。另一方面，也有研究发现，自上而下的信息，比如语言经验，也会影响个体的概率词切分表现（Bonatti et al.，2005；Gómez et al.，2017；Palmer et al.，2019；Poulin-Charronnat et al.，2016；Bosseler et al.，2016）。在词汇知识方面，Palmer 等（2016）在人工语言中嵌入了被试已经掌握的词，比如“tomorrow”，随后，将此条件下被试的测试成绩与不包含被试熟知词条件的成绩进行比较。结果发现，在前一种条件中，被试的迫选正确率显著高于后者，这说明，语言经验（已经掌握的词）能够促进个体在学习阶段切分语流。在词汇的韵律方面，Lew-Williams 和 Saffran（2012）以两组英语母语婴儿为被试，在标准的统计学习任务之前增加了前学习阶段，一组婴儿听一段全部由三音节词组成的人工语言（记为 a），另一组婴儿听一段全部由两音节词组成的人工语言（记为 b）。正式学习材料也是两种，分别由三音节词和两音节词组成（分别记为 A 和 B）。结果发现，前学习阶段听过人工语言 a 的婴儿仅在

人工语言 A 条件下分辨目标词和跨界词，而学习过人工语言 b 的婴儿仅在人工语言 B 条件下能够区分两类选项。这说明，前学习阶段使得婴儿产生了对词长的预期，会引导他们完成统计学习任务。根据自然音布理论，双音节词符合汉语普通话的标准韵律模板，因此普通话母语者有较强的双音节倾向，在日常生活中更偏爱使用双音节词（冯胜利，1998）。比如，当儿童被问到今年几岁的时候，他们普遍会回答"8 岁"，而非"8"。从计量语言学角度看，在《汉语大词典》中，双音节词占比 68.83%，也是不同音节数词中最多的（李斌，刘雪扬，2018）。那么，如果以普通话母语者为实验对象，被试对两音节词的词长期待也会影响他们完成语音统计学习任务，笔者推测学习阶段被试对两音节目标词表征更为牢固，在两音节词迫选试次中回答正确率更高。

转换概率和词长期待两个因素分别属于自下而上和自上而下形式的信息，虽然不乏单独考察某一个因素对统计学习机制影响的研究，但还没有研究在同一个实验内同时设计这两个变量，本研究通过编制两种人工语言来同时考察这两类因素对统计学习机制的影响。为了考察转换概率因素，本研究仿照 Siegelman、Bogaerts 和 Frost 等（2017）的方法设置转换概率不同的两种人工语言作为材料，在高转换概率（high transitional probability，HTP）的人工语言中目标词的转换概率为 1，在低转换概率（low transitional probability，LTP）的人工语言中目标词的转换概率为 0.6。每个人工语言都由两音节词和三音节词组成，前者符合词长期待，后者违背词长期待。实验采用 2（转换概率：HTP、LTP）×2（词长期待：两音节、三音节）的混合实验设计，前者为被试间变量，后者为被试内变量，测试阶段使用的是二选一迫选任务，因变量为被试的正确率。参照以往研究，首先要进行事后检验[4]，通过进行单样本 T 检验比较组平均正确率和随机水平的差异（0.5）来判断在四种条件被试能否完成统计学习任务。在本研究中，还将进行事先对比，相对于事后检验，事先对比（planned contrasts）这种检验方式不仅可以对精细的实验假设进行验证，还可以避免 α 错误的增加，不影响统计检验力。而且，事先对比这一检验思想与传统的事后检验没有必然的联系，即使事后检验不显著，仍旧可以使用事先对比（舒华，张亚旭，2008；Schad et al.，2020；Field et al.，2012）。

基于前言的阐述，低转换概率意味着目标词内部音节间的连结较弱，三音节目标词不符合普通话母语者的词长期待，这两个水平对于被试来说是难度较大的实验条件，由

[4] 一般来说事后检验是在方差分析中对多个水平进行两两比较，但在本研究中，事后检验与事先对比相对，指的是没有提出明确的实验假设后对实验数据进行分析，包括方差分析、T 检验、卡方检验等一系列统计检验方法。

此形成了三个事先假设：(1) 被试在学习 LTP 条件的人工语言中，受词长期待影响，两音节迫选正确率会高于三音节迫选试次；(2) 在违背普通话词长期待的三音节迫选试次中，受到转换概率的影响，在 LTP 条件下被试的正确率会显著低于 HTP 条件下的正确率；(3) 在 LTP 人工语言的三音节迫选试次中，被试的迫选正确率要显著低于在 HTP 人工语言的两音节迫选试次。

2 方法

2.1 被试

64 名南京在校大学生（男 19 名，女 45 名）参加了本次实验，年龄为 18 至 30 岁（$M = 20.91$），均为右利手，非外国语专业，没有接受过正规的音乐训练。所有被试被随机分配到高转换概率组和低转换概率组，其中，高转换概率组 33 人，低转换概率组 31 人。由于 3 名高转换概率组和 1 名低转换概率组被试的实验程序中断，最后两组被试量均为 30 人。被试在实验前了解了实验内容并签订了知情同意书，结束后获得少量报酬。

2.2 材料

在语音统计学习的实验中，要求所使用的词和音节均无任何意义。不同于印欧语言，汉语除轻声以外的所有音节都有声调。参照 Gómez 等（2017）的研究，将所有音节限定为第一声，即选定的音节在第一声时为无意义音节，当换用其他声调时为有意义音节，比如音节“nue1”无意义，但匹配第四声时“nue4”为有意义音节。实验中使用的无意义音节和词见表 1。这样做有三方面考虑，首先，如果每个音节都包括不同声调，那么，音段、声调和音节（音节可以分成音段和声调两个部分，比如音节“chen2”的音段结构为“chen”，声调为第二声）会有不同的转换概率，无形中增加了额外变量。其次，普通话中不同声调音节时长不一，第一声和第三声音节时长偏长，第四声偏短（宋雅男，何伟，2005；冯勇强等，2001），如果将不同声调的音节在时长上进行标准化，听感上会有一些奇怪。最后考虑到如果选用不同声调的音节，那么第三声为曲折调，第一声为平调，无法避免声调调型对统计学习结果的干扰。

表 1 合成人工语言的音节、国际音标以及本实验中的目标词、非词

音节	音标	音节	音标	音节	音标	目标词	非词
nue	nyɛ	rua	ʐua	mei	mei	nueruote	nuegeilai
ruo	ʐuo	dia	tɪa	rou	ʐou	liageirua	liafote

te	tʰɤ	fo	fo	se	sɤ	diafolai	diaruorua
lia	lɪa	lai	laɪ			remei	rerou
gei	kei	re	ʐɤ			rouse	meise

选定音节后，由一名女性普通话母语者在专业录音室进行录音，采样率为 44100Hz。为排除录音者在单独录制每一个音节时产生的重读或停顿现象带来的额外影响，将目标音节放在两个音节之间，并要求录音者一次性录制三个音节，如在音节串 nve1-ruo1-gei1 中，ruo1 为目标音节，所有音节均为第一声。接着，采用 Praat 软件切分出目标音节，并对音节进行标准化，时长 300ms，平均基频 266Hz，音强 70dB，最后通过 Praat 脚本以目标词为单位合成人工语言。

合成目标词后，需要合成转换概率不同的两种人工语言。在 HTP 的人工语言中，每个目标词出现 48 次，共包括 240 个目标词，同时保证同一个目标词不连续出现两次。另外，每个词在人工语言中前半部分和后半部分分布均匀，从而避免某一个目标词在前段或后段多次出现带来首因效应或近因效应，最终 HTP 条件人工语言时长为 3 分 10 秒。在构造 LTP 人工语言时，先合成一个中介人工语言，其中每个目标词出现 80 次，共计 400 词，其他规则与 HTP 一致。然后仿照 Siegelman，Bogaerts 和 Frost（2017）的实验设计，通过将部分目标词替换为填充词来降低目标词的转换概率，LTP 条件下目标词的转换概率为 0.6，对每个目标词需要替换 32 个填充词。构造填充词有三个原则：（1）填充词的所有音节均来自表 1 中的 13 个无意义音节，从而避免填充词的认知加工负荷。（2）对于三音节填充词，保证第一个音节与目标词保持一致，第二个音节有三种可能，分别为另外两个三音节词语的最后一个音节，或者是另外两个两音节词语的最后一个音节；填充词第三个音节可能是一个两音节词的最后一个音节，或者两外两个三音节词的第二个音节。（3）对于两音节填充词，保证第一个音节与目标词第一个音节一致，第二个音节为另外一个两音节词语的最后一个音节，或者是三音节目标词的第二个音节，填充词示例见图 2（上）。在目标词出现次数上，LTP 人工语言与 HTP 条件相同，均出现 48 次，但 LTP 人工语言的转换概率降低到 0.6，其时长为 5 分 10 秒。

在迫选任务中，每个试次的两个选项分别为目标词和非词，其中，非词的每个音节均来源于不同的目标词，非词音节间的转换概率为 0。在每个迫选试次中，两个选项的长度相同，即三音节目标词只和三音节非词进行迫选，两音节目标词只和两音节非词进行迫选。此外，为了避免选项呈现顺序的影响，每个试次会调整选项顺序呈现两次，共计 20 个试次随机播放，三音节迫选 12 个试次，两音节迫选 8 个试次。

实验中的人工语言					
HTP和LTP中的目标词	liageirua	diafolai	nueruote	remei	rouse
LTP中的填充词	lialairuo	diatese	nueruamei	rese	roumei
	liasefo	diaseruo	nuelaigei	refo	rougei
	liatemei	diaruase	nuemeifo	regei	rouruo

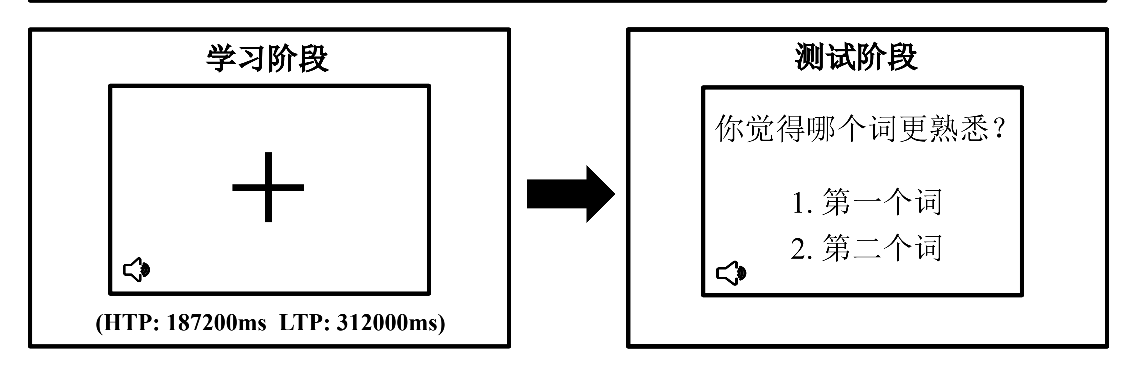


图 2 人工语言填充词示意图（上）和实验程序示意图（下）

2.3 实验程序

实验程序参照经典统计学习范式，分为学习阶段和测试阶段，如图 2（下）。在学习过程开始前，主试讲解实验指导语，先向被试讲解无意义词的概念，随后要求被试认真听人工语言，并提示在人工语言播放之后会要求被试完成测验任务，人工语言播放过程中音量恒定。学习阶段结束后，向被试讲述迫选测验的含义，每个试次会连续播放两个声音刺激，中间有 500ms 间隔，播放之后屏幕会用文字的形式要求被试选择出哪个声音刺激听起来更为熟悉，数字“1”键代表第一个选项，数字“2”键代表第二个选项。正式的迫选试次呈现前，有两个练习试次，由主试协助被试完成。实验全程被试佩戴耳机完成，HTP 条件下大约需要 10 分钟完成实验，LTP 条件下大约需要 15 分钟。

3 结果与分析

3.1 事后检验：四种条件被试正确率检验

对被试在四种条件下（高转换概率两音节词迫选，高转换概率三音节迫选，低转换概率两音节迫选和低转换概率三音节迫选）判断的正确率进行统计，并与随机水平(0.5)进行单样本 T 检验。结果显示，在 HTP 条件中，三音节迫选试次中被试的正确率显著高于随机水平，$n = 30$，$M = 0.61$，$t\,(29) = 3.88$，差值 95%CI = [0.05, 0.17]，$p = 0.001$，$d = 0.71$；两音节迫选试次中被试的正确率显著高于随机水平，$n = 30$，$M = 0.62$，$t\,(29)$

= 3.04，差值 95%*CI* = [0.04, 0.20]，p = 0.005，d = 0.56。最后对 HTP 条件下被试的总体正确率进行检验，发现迫选正确率显著高于随机水平，n = 30，M = 0.61，t (29) = 4.16，差值 95%*CI* = [0.06 0.17]，$p < 0.001$，d = 0.76。

在 LTP 人工语言分析中，发现在两种迫选条件下结果出现差异。三音节迫选试次中被试的迫选正确率与随机水平差异未达到显著水平，n = 30，M = 0.54，t (29) = 1.47，差值 95%*CI* = [–0.01, 0.09]，p = 0.152，d = 0.27；两音节迫选试次中被试的正确率显著高于随机水平，n = 30，M = 0.61，t (29) = 3.41，差值 95%*CI* = [0.05, 0.18]，p = 0.002，d = 0.62。对被试在 LTP 条件下的整体成绩进行检验后发现迫选的正确率显著高于随机水平，n = 30，M = 0.57，t (29) = 2.86，差值 95%*CI* = [0.02, 0.11]，p = 0.008，d = 0.52。

四种条件下被试的正确率如图 3。这些结果说明，虽然被试在学习 HTP 和 LTP 两种人工语言后都可以较好地在迫选任务中正确区分目标词和非词，但在学习目标词内转换概率较低的人工语言过程中，被试对于不符合词长期待的目标词的学习效果较差。

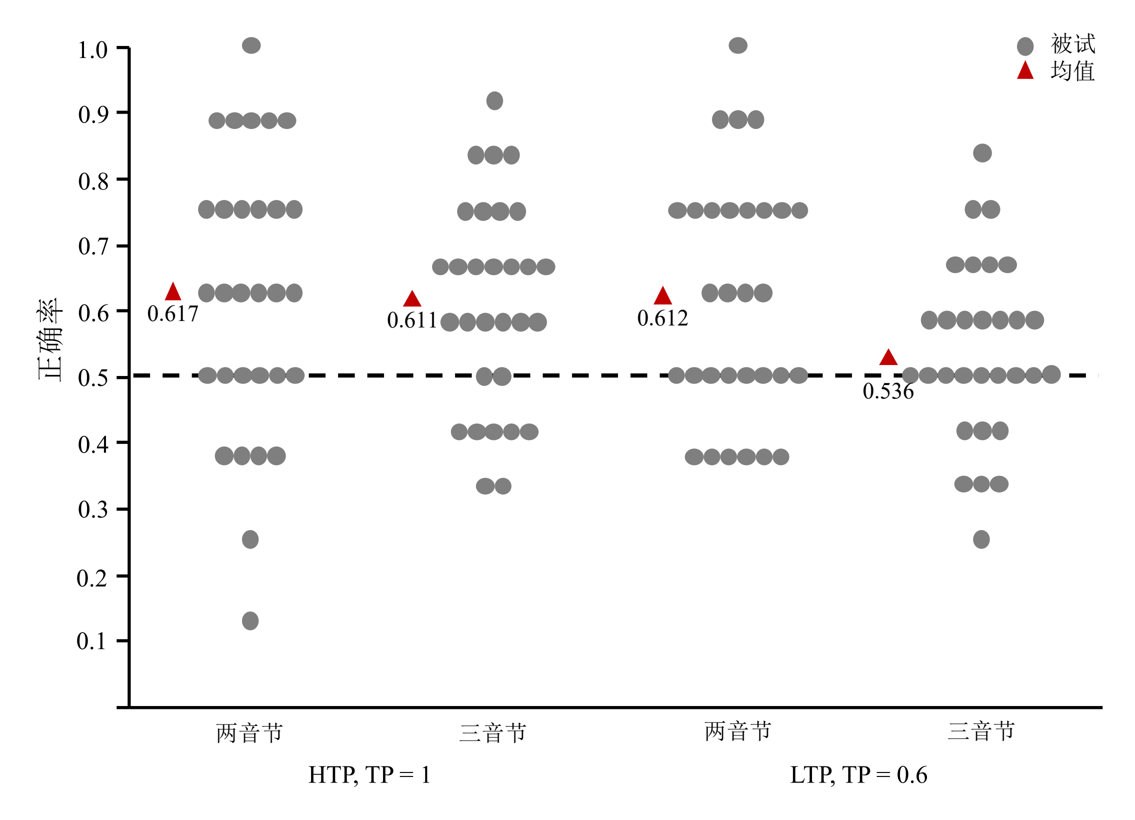


图 3 四种条件下被试迫选测验的正确率

3.2 事后检验：转换概率和词长期待的方差分析

在 R 语言环境下，以转换概率和词长期待作为自变量，以被试迫选测验的正确率为因变量建立线性模型，分别对两个因素的主效应和交互作用进行检验。结果发现，词长期待的主效应未达到显著水平，β = 0.02，t = 1.42，p = 0.157；转换概率的主效应未达

到显著水平，$\beta = -0.02$，$t = -1.37$，$p = 0.172$。转换概率和词长期待的交互作用也没有达到显著水平，$\beta = 0.02$，$t = 1.22$，$p = 0.222$。线性模型参数如表 2。

表 2 转换概率和词长期待对统计学习效果影响的方差分析结果

	自变量	estimate	se	t	p
方差分析	截距	0.59	0.01	41.04	< 0.001**
	TP	–0.02	0.01	–1.37	0.172
	词长期待	0.02	0.01	1.42	0.157
	TP×词长期待	0.02	0.01	1.22	0.222

3.3 事先对比

根据前言所阐述的内容，假设在 LTP 条件下，个体对两音节迫选试次正确率显著高于三音节迫选试次的正确率；在三音节迫选试次中，HTP 条件下的正确率显著高于 LTP 条件；LTP 条件的三音节迫选试次正确率显著低于 HTP 条件的两音节试次迫选正确率。在 R 环境下，使用线性回归通过自定义对比矩阵完成检验。

结果发现，在三音节迫选试次中，被试在 LTP 条件下的正确率显著低于 HTP 条件，$\beta = -0.08$，$t = -2.05$，$p = 0.041$；在 LTP 条件下，被试两音节迫选试次的正确率边缘显著高于三音节迫选试次的正确率，$\beta = 0.08$，$t = 1.87$，$p = 0.062$；最后，被试在学习 HTP 人工语言后对两音节迫选试次的判断情况显著高于被试学习 LTP 人工语言后对三音节迫选试次的判断 $\beta = 0.08$，$t = 1.97$，$p = 0.049$，线性模型结果如表 3。

表 3 事先对比结果

	自变量	estimate	se	t	p
事先对比	截距	0.59	0.01	41.04	< 0.001**
	对比 1	–0.08	0.04	–2.05	0.041*
	对比 2	0.08	0.04	1.87	0.062
	对比 3	0.08	0.04	1.97	0.049*

4 讨论

本研究在经典的统计学习范式下，考察转换概率和被试的词长期待对语音统计学习任务的影响。结果发现，在 HTP 条件下，无论是符合词长期待（两音节迫选）还是不符合词长期待（三音节迫选）的情况，被试都可以较好地分辨目标词和非词，表现出明显

的学习效应；但在 LTP 条件下，被试在不符合词长期待的迫选试次中无法有效分辨两类词语。此外，事先对比结果发现转换概率和词长期待共同影响统计学习效果。

4.1 转换概率和词长期待对统计学习机制的协同影响

转换概率是统计学习领域中的核心概念，其思想渊源最早可以追溯到语言学家哈里斯的音素分布思想（Harris，1954；1955），反映的是相邻音节共现的可能性。由于目标词内音节同时出现且每个音节只出现在一个目标词内，因此对目标词来说，内部两个或三个音节连接紧密，很可能作为一个词被头脑表征，相反，跨界词和非词内的连接强度则逐渐降低，因此被以词为单位表征的可能性较小。基于这样的逻辑关系，大量统计学习研究都证实，个体可以根据音节间的转换概率对语流进行切分，但以往研究中迫选试次呈现的两个选项不仅在转换概率上存在差异，在人工语言中出现的频次也存在差异（Aslin et al.，1998；Thiessen & Saffran，2003；Shoaib et al.，2018）。比如，在人工语言 ABC-DEF-ABC-GHI……中，虽然目标词 ABC 内的转换概率高于跨界词 CDE，但音节 C 后会出现三种目标词（人工语言由四个目标词合成），因此跨界词 CDE 在人工语言中出现的频次仅为目标词 ABC 的三分之一，那么，即使实验结果发现被试能够在迫选任务中区分目标词和跨界词，也无法确定是因为两类词的转换概率不同还是频次不同。更为重要的是，已有研究发现，在语言中词频是影响词切分和词语习得的因素（Frost et al.，2019）。笔者通过改进实验范式实现了考察不同转换概率是否会影响统计学习的目标。在设计不同转换概率人工语言的基础上，通过增加填充词保证了目标词出现在两种人工语言中的频次相同（均为 48 次），而且高低转换概率条件下的目标词均与词长相同的非词进行迫选，其中，非词内音节的转换概率为 0，在人工语言中出现的频次也为 0，因此在这两个条件下实验结果的任何差异仅仅来自于不同的转换概率。本实验中虽然转换概率的主效应未达到显著水平，但事先对比结果发现对于三音节迫选试次来说，被试在高转换概率的条件下的学习效果更好，这提示转换概率对统计学习效果的影响受限于迫选任务的难度。

本研究同时关注词长期待对统计学习的影响。从计量语言学角度来看两音节词在汉语词汇中占据主导地位，而且有学者指出，普通话母语者对汉语词长的期待是一种主观的、韵律上的预期。冯胜利（1996）明确指出，“两音节”更容易被感知成词，而超过两音节的组合则不易被感知为词，甚至在句法上被判断为不是词的词组，如“踢球”，也常被母语者认为是词，这体现了韵律对句法的压制（冯胜利，1996；张辉等，2013）。在本研究中，发现被试在切分 LTP 条件下的人工语言过程中，对两音节迫选试次的判断效果

明显好于三音节迫选试次，这说明普通话母语者对词长的韵律期待在一定程度上会引导统计学习过程，尤其是在测验任务难度较大时词长期待才会体现出实验效应。同时，对被试在四种条件下的迫选正确率进行了单样本 T 检验发现，只有在完成 LTP 条件下的三音节迫选试次时，被试无法区分目标词和非词。在这种条件下，三音节词不符合被试的词长期待，并且三音节目标词的转换概率较低，从而导致学习效果的下降。总而言之，事先对比和事后检验均证明两个因素协同影响统计学习的过程。

需要指出的是，在两种词长期待对应的迫选试次中，两音节词中仅包含一个转换概率，而三音节词中包含两个转换概率，虽然两类词在记忆资源的需求上有所不同，但笔者认为记忆负担不是导致 LTP 条件下三音节迫选试次正确率显著低于两音节迫选试次的正确率的主要原因。如果存在记忆负担因素的影响，那么，在 HTP 条件下三音节迫选试次正确率和两音节迫选试次正确率也应该存在显著差异，但实验结果显示在 HTP 条件下三音节迫选成绩与两音节迫选成绩没有显著差异。本研究的方差分析结果显示，两个因素的主效应以及交互作用均未达到显著水平，尤其是 HTP 条件下被试正确率为 0.614，LTP 条件下的正确率为 0.574。这与 Bogaerts 等（2016）实验发现被试的迫选正确率随着 TP 的上升而显著上升有所不同，可能的原因是听觉通路比视觉通路更加复杂，信息视觉输入和信息听觉输入的过程不同，在听觉感受器和大脑皮层之间存在着更多的环节（Belliveau et al.，1991；Hudspeth，1989），那么，以相同的转换概率合成两种模态下的实验材料，还应该考虑材料呈现时间和学习时长等因素。

4.2 复杂词长条件下的语音统计学习研究

在以往研究中，研究者在合成人工语言过程中普遍都使用同一长度的词，比如三音节（Antovich & Estes, 2017; Estes, Gluck, & Bastos, 2015; Estes & Lew-Williams, 2015）或两音节（Mirman et al., 2008；Gómez et al.，2017），而本研究则选择使用两种长度的无意义词合成人工语言。有研究指出，被试在完成统计学习任务过程中注意资源起到重要作用（Toro et al.，2005），由于人工语言中本身所有音节时长固定，再加上无意义词语长度相等，那么随着学习过程的深入，根据动态注意理论（Jones & Boltz，1989），被试会形成固定的节奏预期，从而提高被试的学习效果。Hoch 等（2013）通过对比被试学习等长词（六个三音节词）合成的人工语言和不等长词（三个三音节词和三个两音节词）组成的人工语言后的切分效果，发现前一种条件下被试的正确率显著高于后一种情况；另外，Johnson 和 Tyler（2010）也发现当目标词包含两种长度时，在再认过程中婴儿无法分辨目标词和跨界词，这两个研究都说明了在统计学习研究中应该考虑目标词长度一

致性对学习效果的影响。本研究中的人工语言由三个三音节和两个两音节合成，不仅还原了最为纯粹的语音统计学习过程，还实现了对普通话母语者词长期待因素的考察。

4.3 转换概率和词长期待对统计学习机制的影响进程

早期研究者多关注个体能否利用语言中的概率信息这一自下而上形式的信息来实现词切分和语言学习，却很少关心影响因素的作用发生在统计学习机制的哪些阶段。统计学习范式采用的是“学习-再认”这一经典的线下范式，对于完成迫选任务的成人被试来说，每一个目标词和跨界词都会多次出现在迫选任务中，因此被试的迫选结果不仅来自于学习过程中的学习效果还来自于迫选任务中的二次学习效果（Siegelman，Bogaerts，& Frost，2017；Siegelman，Bogaerts，& Christiansen et al.，2017）。研究者进一步指出统计学习效果在学习阶段和测试阶段均有体现，在学习阶段可以分为对输入信息的感知编码（对听觉音节和视觉图片的初步表征）、对分布信息的识别（对转换概率信息的加工）和对组合单元的记忆（将切分出来的音节组合储存在短时记忆中），在测试阶段表现为迫选任务中的二次学习（Siegelman，Bogaerts，& Christiansen et al.，2017）。在本研究中，对于学习 LTP 人工语言的被试来说，虽然他们在完成两种迫选试次的正确情况出现显著差异，但由于在每个迫选对中，两个选项的长度相同，均为两音节或三音节，因此正确率上的差异是来自于学习过程中对两类词的切分效果。换句话说，虽然本研究设计了自下而上和自上而下的两种信息，但他们都是在被试学习人工语言过程中引导被试切分语流。我们推测词长期待不太可能影响头脑对输入信息的编码以及概率信息的识别，更可能干扰将切分出的三音节目标词保存在记忆中的过程，即符合词长期待的两音节目标词具有更高的保存优先级。不过，不同语言经验对统计学习影响的时间进程可能有所不同。Toro 等（2011）采用线下范式发现语音经验（音系规则）并不影响个体在学习过程利用概率信息切分语流，反倒是影响在迫选任务中的再认。除了词长期待以外，也有学者关注其他语言经验对统计学习效果的影响，比如前言提到的词汇知识（Palmer et al.，2019；Poulin-Charronnat et al.，2016）、儿向语的语气和语调（Thiessen et al.，2005）等。今后的研究除了探讨多种信息对统计学习效果的影响外，还应该关注不同信息对统计学习机制影响的进程和强度。

5 结论

转换概率是统计学习中的核心概念，两音节词长期待是普通话母语者典型的语言经

验，本研究考察这两个因素是否会影响个体的语音统计学习效果，结果发现转换概率和和被试的词长期待协同影响语音统计学习的效果，体现了自下而上信息和自上而下信息的整合。

参考文献

Antovich, D. M., & Estes, K.G. (2017). Learning across languages: bilingual experience supports dual language statistical word segmentation. *Developmental Science*, *21*(2). doi: 10.1111/desc.12548

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive science*, *36*(2), 286–304.

Bosseler, A. N., Teinonen, T., Tervaniemi, M., & Huotilainen, M. (2016). Infant directed speech enhances statistical learning in newborn infants: An ERP Study. *PLoS ONE*, 11(9). doi: 10.1371/journal.pone.0162177

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. J*ournal of Memory and Language*, *83*, 62–78.

Belliveau, J. W., Kennedy, D. N., Mckinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., ... & Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, *254*(5032), 716–719.

Bogaerts, L., Siegelman, N., & Frost, R. (2016). Splitting the variance of statistical learning performance: A parametric investigation of exposure duration and transitional probabilities. *Psychonomic Bulletin & Review*, *23*(4), 1250–1256.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science*, *16*(6), 451–459.

Emberson, L. L., Misyak, J. B., Schwade, J. A., Christiansen, M. H., & Goldstein, M. H. (2019). Comparing statistical learning across perceptual modalities in infancy: an investigation of underlying learning mechanism(s). *Developmental Science, 22*(6). doi: 10.1111/DESC.12847

Erickson, L. C., Thiessen, E. D., & Estes, K. G. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language, 72*, 49–58.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*(3), 254–260.

Estes, K. G., Gluck, C. W., & Bastos, C. (2015). Flexibility in statistical word segmentation: finding words in foreign speech. *Language Learning and Development*, *11*(3), 252–269.

Estes, K. G., & Lew-Williams, C. (2015). Listening through voices: infant statistical word segmentation across multiple speakers. *Developmental Psychology*, *51*(11), 1517–1528.

Feng, S. (1998). On the "natural foot" of Chinese Mandarin. *Studies of the Chinese Language, 1*, 40–47.

[冯胜利. (1998). 论汉语的“自然音步”. *中国语文*, (1), 40–47.]

Feng, S. (1996). Prosodic structures in Mandarin and its restrictions on syntactic structures. *Linguistic Studies*, (1),108–122.

[冯胜利. (1996). 论汉语的韵律结构及其对句法构造的制约. *语言研究*, (1),108–122.]

Feng, Y. Q., Chu. M., He., L., Lv., S. N. (2001). Statistical analysis of syllable duration in Chinese discourse. Modern phonetics in the new century: Proceedings of the 5th National Conference on modern phonetics.

[冯勇强, 初敏, 贺琳, 吕士楠. (2001). 汉语话语音节时长统计分析. 新世纪的现代语音学——第五届全国现代语音学学术会议论文集.]

Field, A., Miles, J., & Field., Zoe. (2012). *Discovering statistics using R*. London: SAGE. Publications Ltd.

Frost, R., Armstrong. B. C., & Christiansen. M. H. (2020). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125.

Frost, R. L., Monaghan, P., & Christiansen, M. H. (2019). Mark my words: High frequency marker words impact early stages of language learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *45*(10), 1883–1898.

Gómez DM, Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2017). Statistical speech segmentation in tone languages: the role of lexical tones. *Language & Speech*, *61*(1), 84–96.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162.

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*(2), 190–222.

Hoch, L., Tyler, M. D., & Tillmann, B. (2013). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychonomic Bulletin & Review*, *20*(1), 142–147.

Hudspeth, A. J. (1989). How the ear's works work. *Nature*, *341*(6241), 397–404.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Development Science*, *13*(2), 339–345.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: expectations about word length guide infant statistical learning. *Cognition*, *122*(2), 241–246.

Li, B., & Liu, X. (2018). A quantitative analysis of Chinese vocabulary evolution based on The Great Chinese Dictionary. *Journal of Nanjing Normal University (Social Science Edition)*, *5*, 152–160.

[李斌，刘雪扬. (2018). 基于《汉语大词典》的汉语词汇历时演变计量研究. *南京师大学报(社会科学版)*, (05), 152–160.]

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280.

Onnis, L., & Thiessen, E. D. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268–284.

Palmer, S. D., Hutson, J., White, L., & Mattys, S. L. (2019). Lexical knowledge boosts statistically-driven speech segmentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 139–146.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology*, *69*(12), 2390–2401.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: eight-month-old infants track backward transitional probabilities. *Cognition, 113*(2), 244–247.

Potter, C. E., Wang, T., & Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive science, 41*(S4), 913–927.

Poulin-Charronnat, Bénédicte, Perruchet, P., Tillmann, B., & Peereman, R. (2016). Familiar units prevail over statistical cues in word segmentation. *Psychological Research*, *81*, 990–1003.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developmental Science*, *21*(4). doi: 10.1111/desc.12593

Revelle, W. (2016). *psych: procedures for personality and psychological research.* R package version 1.6.6. Available at: http://cran.r-project.org/package=psych.

Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: the role of tonal structure. *Developmental Science*, *6*(1), 35–43.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*(1), 181–203.

Saksida, A., Langus, A., & Nespor, M. (2017). Co-Occurrence statistics as a language-dependent cue for speech segmentation. *Developmental Science*, 20(3). doi: 10.1111/desc.12390

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*. doi: 10.1016/j.jml.2019.104038

Shoaib, A., Wang, T., Hay, J. F., & Lany, J. (2018). Do infants learn words from statistics? Evidence from English-learning infants hearing Italian. *Cognitive Science*, *42*(8), 3083–3099.

Shu, H., & Zhang, Y. X. (2008). *Research methods in psychology: Experimental design and data analysis*. Beijing, China: People's Education Press.

[舒华, 张亚旭. (2008). *心理学研究方法:实验设计和数据分析*. 北京：人民教育出版社.]

Song, Y.N. & He, W. (2005). Statistical analysis of monosyllabic duration in standard Mandarin. *National Conference on man machine voice communication*.

[宋雅男, 何伟. (2005). 标准普通话单音节时长统计分析. 全国人机语音通讯学术会议.]

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42* (S3), 692–727.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B*. doi: 10.1098/rstb.2016.0059

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, *64*(2), 171–180.

Toro, J. M., Sinnett, S., & Sotofaraco, S. (2005). Speech segmentation by statistical learning depends on attention. Cognition, 97(2). doi: 10.1016/j.cognition.2005.01.006

Wang, T. L., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.00953

Yu, W., & Liang, D. (2018). Word segmentation cues in the process of spoken language. *Advances in Psychological Science*, *26*(10), 1765–1774.

[于文勃, 梁丹丹. (2018). 口语加工中的词语切分线索. *心理科学进展*, *26*(10), 1765−1774.]

Zhang, H., Sun, H., & Gu, J. (2013). The Interaction between prosody and syntax in Chinese idiom processing: An event-related potentials investigation. *Journal of Foreign Languages*, (1), 22–31.

[张辉, 孙和涛, 顾介鑫. (2013). 成语加工中韵律与句法互动的事件相关电位研究. *外国语: 上海外国语大学学报*, (1), 22−31.]

Transitional probabilities and expectation for word length impact verbal statistical learning

YU Wenbo[1], WANG Lu[1], QU Xingfang[1], WANG Tianlin[2], ZHANG Jingjing[3], LIANG Dandan[1]

([1]School of Chinese Language and Culture, Nanjing Normal University, Nanjing 210097, China)

([2]School of Education, University at Albany, State University of New York, New York 12222, USA)

([3]School of Psychology, Nanjing Normal University, Nanjing 210097, China)

Abstract Statistical Learning (SL) has long been established as a powerful mechanism in language learning and development (Saffan et al., 1996). Within this framework, transitional probability (TP) of various levels have been shown to confer differing task performance for adults (Wang & Saffran, 2014; Mirman et al., 2008). Recent studies have also highlighted the role of linguistic experience in the process of SL (Frost et al., 2020; Siegelman et al., 2018). However, it remains unclear whether different word lengths as well as varying levels of TPs may impact the segmentation of continuous speech.

Method: Sixty native Mandarin monolinguals participated in a word segmentation task. An artificial language was designed with the same flat tone paired with 13 syllables, resulting in two disyllabic and three trisyllabic monotonic words. While only the segmental tier offered reliable information to segmentation, information from the suprasegmental level ensured that each word was phonologically legal in Mandarin. The words were then combined into two conditions of a monotonic artificial language: for the hTP language, all TPs within words were 1.0; whereas they were 0.6 in the lTP condition. Two types of nonwords (trisyllabic and disyllabic) were created for the test phase, then paired with target words of equal length in each trial. Adults were first exposed to the monotonic artificial language and then tested in a 2 alternative forced-choice task (2AFC) to decide whether a word or a nonword sounded more familiar.

Results: The mixed two-way ANOVA with word length (disyllables vs. trisyllables) as a within-subject factor and TP (high TP vs. low TP) as a between-subjects factor yielded nonsignificant effects for either word length or TP levels. There was also no significant interaction. A series of one-sample t-tests were conducted between the participants' average accuracy and chance level (.5). Participants preferred words over nonwords in most conditions except for trisyllabic contrasts in the low TP condition. Additional planned contrasts among the conditions revealed that when the TPs were low, participants performed significantly better in the disyllabic condition than in the trisyllabic condition; under the trisyllabic condition, high TP yielded better performance than low TP.

Discussion: In the low TP condition, the superior outcome of disyllabic contrasts might stem from the Mandarin speakers' prior linguistic experience—their expectation that words should be of two syllables. For the trisyllabic contrasts, lower TPs may provide relatively weakened statistical regularities for tracking word boundaries, which may in turn lead to difficulty extracting words. Importantly, our findings show that when both factors present difficulties (e.g., trisyllabic contrasts in the low TP condition), such that the word length violates the listeners' expectation and the TPs do not provide high levels of consistency, word segmentation can no longer be supported. The current study showed for the first time that when combining TP and word length in the speech

material, these two factors impact word segmentation in a complex manner. This study offers new insight for future SL designs as well as potentially informative directions in exploring how individual differences based on linguistic backgrounds may manifest itself in word segmentation tasks.

Key words verbal statistical learning，expectation for word length，transitional probabilities，prosody